

第5章 体系の異なる分類の対応関係と変換 - グループ化および切断による商品分類の試み -

権利	Copyrights 日本貿易振興機構（ジェトロ）アジア経済研究所 / Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO) http://www.ide.go.jp
シリーズタイトル(英)	I.D.E. statistical data series
シリーズ番号	67
journal or publication title	World trade data system : arrangement and its application
page range	69-82
year	1995
URL	http://hdl.handle.net/2344/00009161

第5章

体系の異なる分類の対応関係と変換 —— グループ化および切断による商品分類の変換の試み ——

野田 容助
山本 泰子

はじめに

国際間の合意にもとづいて1988年以降適用可能な国からHS準拠で貿易商品进行分类する国が多くなるにしたがって、国際機関が提供する貿易統計データもSITC・R2からHSに対応するSITC・R3によることが多くなってきた。これにともない時系列的なデータの接続という意味においてSITC・R2からSITC・R3、または、その逆の変換が必要になってきた。

本章では一般に体系の異なる分類コードAと分類コードBの間に対応関係がある時、分類コードAの統計値を分類コードBへ変換させる方法を検討し、この方法をSITC・R2とSITC・R3の対応に適用することを目的とする。

体系の異なる分類どうしを結び付けるには両者の対応関係を明らかにした対応関係コード表が必要であり、対応関係コード表を使用する場合には2つの分類どうしがどのような対応関係にあるかを検討することが重要な問題になる。分類Aと分類Bの対応関係コード表を考えると、この対応関係コード表の中で分類の核になる閉じた対応関係にある分類コードの集まりを「グループ」ということにする。グループとは第7章で紹介される2つの分類から「共通に導出可能な最も詳細な分類（FCD:Finest Common Derivative）」に対応する分類である（注1）。すなわち、対応関係に少なくとも1つの共通した結合があればつないでいき、結合が無くなったところまでの構成要素でグループを決めるという方法を用いる。形式的にグループを作るとグループとしての共通性が問題になるが、この点については以下で述べる切断という方法でサブグループを作り、共通した特性を持たせるように配慮している。。

一般的に対応関係のグループは次のような4つのタイプに分けることができる。タイプ1は分類Aと分類Bが1対1の対応関係である。このタイプではグループに含まれる分類コードの個数は1個である。タイプ2は分類Aと分類Bが1対多の対応関係である。グループに含まれる分類コードの個数は分

類Bに含まれる分類コードの個数に等しい。タイプ3は分類Aと分類Bが多対1の対応関係であり、タイプ2とは逆にグループに含まれる分類コードの個数は分類Aに含まれる分類コードの個数に等しい。タイプ4は分類Aと分類Bが多対多の対応関係である。このタイプのグループに含まれる分類コードの個数について特に決まったパターンはない。

このようにして分類された4つのタイプの対応関係の中で、分類Aのコードで表された統計値が対応する分類Bのコードではどのように推計されて変換されるかが本章の検討課題であるが、個別の分類コードについて考えたときには、4種類の対応関係の中でタイプ1とタイプ3のグループは分類のコードAの統計値を直接分類B対応させるかあるいは統合するだけで変換のための処理をすることができる。これに対して、タイプ2とタイプ4のグループはある条件のもとで分類Aのいくつかのコードの統計値を分類Bのコードへの配分および統合という変換のための処理が必要になる。特に配分については配分のための推計方法も一般的な手法としては存在せず、また、配分による推計は必ずしも一意的な解が存在するとは限らない。そのため個別の分類コードを対象とした変換はいろいろと問題があると考えられる。

それに対して、グループごとに統計値を対応させればおのおの分類コードの統計値を配分する必要がなくなり、配分による推計誤差が無くなる。したがって本章では統計値の変換はグループごとにおこなうという方法を採用することにする。すなわち、分類間の対応関係を見る場合に個別の分類コードを考えずに関係ある個別分類コードの閉じた集まりであるグループを1つの対応関係とするということである。

グループを1つの対応関係と見る方法はグループの要素の数が少なければそのグループが持つ特性を明らかにできるが、要素の数が大きくなるとそれが持つ特性の範囲が広くなりすぎて共通性を持つ要素の特性が薄れ、内容がわからなくなってしまう恐れがでてくる。したがって、グループ化を考えるさい

にグループごとの対応関係を明らかにすると同時にグループが大きくなったときの取り扱いとして小さなまとまりとしてのサブグループの必要が生じてくる。

グループからある対応関係を取り除いた残りの対応関係がいくつかのサブグループに分かれるとき、グループが切断されたといいそのとき取り除いた対応関係を切断の要素という。サブグループ化の方法はいくつかの対応関係が切断された対応関係コード表に対して再度グループ化とタイプ分けを試みるということである。また、切断というのは対応関係コード表の特性を決めることなので、切断によるサブグループ化は対応関係コード表に対する1つのモデルであると考えることができる。この方法の適用例として本章では商品分類に関する切断なしの基本モデルGRT₃₂とアジア経済研究所作成の独自の切断モデルGRT₃₂(IDE)を考える。

切断についてはいろいろな方法が考えられるが、本章ではグループ内に存在するすべての対応関係の内容を個別に検討して比較的關係がなさそうだと判断される対応関係を再調整するという方法を用いることにする。GRT₃₂(IDE)の作成にさいしては切断のために個別商品の特徴に応じて対応関係の類型化をおこない、具体的な切断方法を明らかにする。GRT₃₂(IDE)の対応関係は「GRT₃₂(IDE)およびアジア研統一国コード表」で示されている。

第1節 対応関係のグループ化とそのタイプ分け

体系の異なる分類コードAと分類コードBの間に対応関係コード表があるとする。対応関係コード表の中で対応する分類Aと分類Bの分類コードが互いに結びついている箇所をそれぞれ関連させて閉じた対応関係を作るのがグループ化である（補論の「A. グループ化の方法について」、「B. グループ化における対応関係の関連情報」を参照のこと）。このようにして得られた個々のグループに対して対応関係のタイプを4つに分類することができる。タイプ1からタイプ3までの対応関係はお互いの関係が容易に把握できるが、タイプ4のグループ分けはそれほど簡単ではない。

(1) タイプ1は分類コードAの1個に対して分類コードBが1個という1対1の対応関係であり、分類AとBの分類コードの結合の手がともに1本ずつであるグループである。1個の対応関係で閉じているのでグループに含まれる分類コードの個数は1個である。

(2) タイプ2は分類Aの分類コード1個に対して分類Bの分類コードが複数個という1対多の対応関係である。この対応関係は分類Aに属する1個の分類コードが分類Bに対応する個数分の結合の手を持つものに対して、分類Bに属する複数個の分類コードは1本ずつの結合の手を持つ対応関係の集まりである。表1にはあるグループの対応関係の状態とその時の結合の手の数が見られている。ここで、コードを表現するとき、分類体系とコードを合わせて、

「分類体系：コード」

と表すことにする。分類Aのコードである12111はA:12111と表される。A:12111はB:12111とB:12119への2本の結合の手を持つ。結合の手は分類コードAの出現頻度に等しいので、Aの頻度は2である。B:12111とB:12119は結合の手はともに1本であるので、該当するBの頻度の項はそれぞれ1となる。

表1 タイプ2の対応関係とその対応関係コード

対応関係 A B	対応関係コード A B	頻度 A B
12111: < 12111	12111 12111	2 1
12119	12111 12119	2 1

(出所) 著者作成

タイプ2は分類Bの頻度がすべて1である分類コードの合計と1個しかない分類Aの分類コードの頻度が等しいグループである。この例の場合には、分類Bの頻度の合計は2であり、分類Aの頻度は2なのでグループ内の対応関係の妥当性が示される。グループに含まれる分類コードの個数は分類コードBの個数に等しい。

(3) タイプ3は分類コードAの複数個に対して分類コードBが1個という多対1の対応関係である。この対応関係は分類Aに属する複数個ある分類コードが1本ずつの結合の手を持つものに対して、分類Bに属する1個の分類コードは分類Aに対応する個数分の結合の手を持つコンバータの集まりである。タイプ2とは逆に分類Aの頻度がすべて1である分類コードの合計と1個しかない分類Bの分類コードの頻度が等しいグループである。グループに含まれる分類コードの個数は分類コードAの個数に等しい。

(4) タイプ4は分類コードAの複数個に対して分類コードBが複数個という多対多の対応関係を表し、(1)から(3)以外のタイプの集まりである(注2)。表2によればA:00122は結合の手を1本持っており、該当するコードAの頻度は1である。

A:01112は結合の手を2本持っているため、A:01112の頻度は2である。同じようにB:00122は2本、B:01112は1本の結合の手を持っているので、それぞれのコードの頻度は2、1となる。

表2 タイプ4の対応関係とその対応関係コード

対応関係		対応関係コード		頻度	
A	B	A	B	A	B
00122	:00122 01112	00122	00122	1	2
01112		01112	00122	2	2
		01112	01112	2	1

(出所) 著者作成

タイプ4ではグループ内の分類AとBの対応関係はグループ内にある分類Aのコードの頻度の合計と分類Bの頻度の合計が等しくなる。この例の場合には、グループ内に表れる分類Aのコードは00122と01112の2つがあり、それらのAの頻度はそれぞれ1と2であるため、分類Aの頻度の合計は3となる。また、分類Bのコードは00122と01112の2つがあり、それらのBの頻度はそれぞれ2と1であるため、分類Bの頻度の合計は3となる。したがって、分類AとBのそれぞれの頻度の合計はともに3になり、対応関係の妥当性が示される。グループに含まれる分類コードの個数はできあがったグループの数に等しい。

第2節 グループおよびサブグループによる分類間の変換

体系の異なる分類間における対応関係のグループ化とそのタイプ分けが明らかになれば次は本章の中心課題である分類間の変換の問題である。個別の分類コード間の変換では4種類の対応関係が存在し、この中でタイプ1とタイプ3は分類のコードAの統計値を直接分類Bに対応させるかあるいは統合するだけで変換のための処理をすることができる。これに対して、タイプ2とタイプ4はある条件のもとで分類Aのいくつかのコードの統計値を分類Bのコードへの配分という処理が必要になる。配分についてはいろいろとやっかいな推計の問題が生じるほか、推計誤差も無視できないと考えられる。したがって個別の分類コードを変換の対象とせずにグループを1つの対応関係とするのはごく自然な発想である。対応関係のグループごとに分類コードAの統計値を分類コードBへ変換させると対応関係のタイプにかかわらず直接に分類間の変換がおこなえる。すなわ

ち、分類間の変換をする場合に個別の分類コード間の対応関係を対象とせずにグループを1つの対応関係とするということである。タイプ3の場合にも正確に言えばグループという概念を取り入れたときのみ統合が可能なのである。本章では分類間の変換にさいしてはこの方法を採用することにする。

2-1 切断によるサブグループ化

グループを1つの対応関係とする方法ではグループの要素の数が少なければそのグループが持つ特性を明らかにできるが、要素の数が大きくなるとそのグループの持つ特性の範囲が広くなりすぎて内容がわからなくなってしまう恐れがでてくる。したがって、大きなグループをいくつかの比較的小さなサブグループに分けることができれば解釈がしやすいわけである。ここでいうサブグループとは厳密な意味におけるグループ内の分割ではなくて互いに少数の共通の分類コードが含まれていても類似のものがまとまっているという意味でのゆるやかな対応関係の集まりをさす。したがって、サブグループは一意的には決まらない。

タイプ4のグループの中で2つのサブグループが1つの対応関係でつながっている場合を想定する。サブグループを作成する方法として、この2つのグループを結ぶ関係を取り除けば両者はそれぞれが固有のグループを構成できるという考えを利用する。ある1つの対応関係を取り除くとグループが2つに分かれるとき、この対応関係によってグループが「切断」されたといいそのときに取り除いた対応関係を「切断の要素」ということにする。切断の要素の数が複数個のときはサブグループも複数個できることがある。

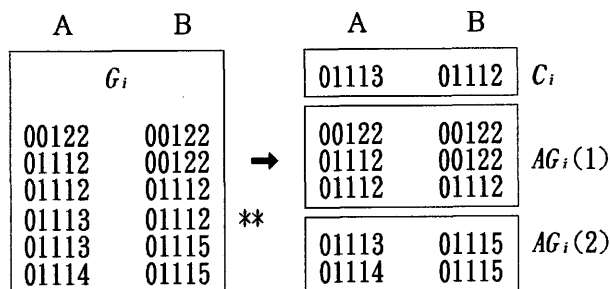
グループの*i*番目を*G_i*、切断の要素の集まりを*C_i*で表す。グループから切断の要素を取り除いた*G_i-C_i*に対してグループ化をおこないサブグループを作る。*n*個のサブグループを*SG_i(j)* *j*=1…*n*とするとグループ*G_i*は、

$$G_i = \{SG_i(1) \cup \cdots \cup SG_i(n) \cup C_i\}$$

と分割される。

図1にグループとサブグループの関係の例が示されている。この例は*n*=2の場合である。グループ*G_i*の中から「**」で示された対応関係A:01113とB:01112の結びつきを切断の要素として取り除き*C_i*としたとする。*G_i*から点線を取り除いた*G_i-C_i*をグループ化すると*AG_i(1)*と*AG_i(2)*の2つのサブグループが得られる。

図1 グループとサブグループの関係



(出所) 著者作成

サブグループがグループから切断の要素を取り除いた対応関係コード表に対して再度グループ化をすることで得られるということは、切断の仕方によってサブグループの内容や個数が決まるということである。このことにより、切断というのは対応関係コード表のグループ化に対する1つのモデルと考えることができる。次節で紹介する商品分類に関する切断なしの基本モデルGRT₃₂やアジア経済研究所作成による独自の切断モデルGRT₃₂(IDE)はグループ化に対するモデルである。

2-2 切断の方法

切断の仕方にはいろいろな方法が考えられる。例えば、結合の手の数が多い対応関係は分類を複雑にする原因であると想定されるのでこの対応関係を順次切断する候補に取り入れてその中から切断の対応関係を探し出す方法や少数個のすべての組み合わせを切断の候補にしてその中から切断の対応関係を探し出す機械的な処理方法等が考えられる(注3)。

本章では機械的に切断する方法を取らずにグループ間の分類の内容を個別に検討して比較的關係がなさそうだと判断される対応関係を切断するという方法を採用することにする。この方法ではまず最初グループの中の対応関係の中で比較的關係がないと考えられる対応関係を切断の候補として G_i の中から取り除くことから始める。この段階ではグループの要素数が多い場合には取り除いた対応関係が切断かどうかを判定するのは難しく、ただ切断の候補を決めるだけである。対応関係が切断かどうかの判定は後でまとめておこなう。

C_i の要素を決めるために、 G_i から取り除いた切断の候補の対応関係の集まりを C_i^* とする。 C_i^* に属している対応関係が切断であるかどうかを確認するため、 C_i^* に属するすべての対応関係を1つずつ検査して以下のような状態に応じて C_i かサブグループ $SG_i(j)$ $j=1\cdots n$ のどちらかに割り振りグループを再調整をする。

(1) 検査する対応関係のAとBの分類コードがそれぞれ異なるサブグループに属していれば G_i-C_i にこの対応関係を追加すれば再グループ化によって異なるサブグループが1つにまとまりサブグループの数が1つ減少する。すなわち、この対応関係を取り除くことによって1つのサブグループが2つに分かれる。この対応関係は切断であるので C_i のメンバーとする。

(2) 対応関係のAとBの分類コードがともに同じサブグループ $SG_i(j)$ に属していれば G_i-C_i にこの対応関係を追加すれば再グループ化によってサブグループの個数に増減が生じない。この対応関係は切断ではないので $SG_i(j)$ のメンバーとする。

(3) 対応関係のAとBの分類コードのどちらか一方のみがサブグループ $SG_i(j)$ に属していなければ再グループ化によってサブグループの個数が変化しない。この対応関係は切断ではないので C_i のメンバーではない。しかしながらこの対応関係をどちらかが属しているサブグループ $SG_i(j)$ に加えるといままどここのサブグループにも属していなかった分類コードがこのサブグループのメンバーになる。この対応関係を追加する前後で割り振りの基準が変わってしまう。さらに、付け加える C_i^* の対応関係の並べ方によっても基準が変わり(1)になるか(2)になるのか判定結果が一意的でなくなる。したがって、(3)の場合には切断ではないが C_i のメンバーとする。

(4) 対応関係のAまたはBの分類コードのどちらもサブグループに属していなければ G_i-C_i にこの対応関係を追加すれば再グループ化によってグループが増加する。この対応関係によって切断とは無関係に新たにグループができるが、(3)と同じようにこの対応関係を追加する前後で新しいグループの要素が変わってしまい、しかもその結果は一意的でない。したがって、切断ではないが C_i のメンバーとする。

この方法を採用するには対象としている分類に関する詳細な知識がないと対応関係間の切断がうまくいかない。また、グループに含まれる分類コードの数が少ない場合にはこの方法を用いるのが確実であるが、分類コードの数が多くなれば分類の内容を検討するのに膨大な労力が必要になることを覚悟しなければならない。サブグループ化は本来ならばつながっているはずの対応関係を恣意的に断ち切って作るのでその対応するサブグループ間の分類の内容については十分に吟味することが必要である。また、切断によるサブグループ化は基本モデルGRT₃₂の対応関係を変えることになりそれにとまって変換

による誤差が生じるので利用のさいには注意する必要がある。このとき得られたサブグループはまた4つのタイプに分けることができる(補論の「C. サブグループ化における対応関係の関連情報」を参照のこと)。

第3節 商品分類SITC・R3とSITC・R2の対応関係

商品分類SITC・R3とSITC・R2の対応関係コード表はUN統計局発行の「Standard International Trade Classification Revision 3(Statistical Papers Series M no.34/Rev3, United Nations 1986)」から得られた対応関係を基本とする。切断なしの対応関係から構成されるこの対応関係コード表を基本モデルとしてGRT₃₂とする。GRT₃₂は2,654個のSITC・R3の個別分類コードと1,831個のSITC・R2の個別分類コードの間のいくつかの組み合わせにより5,245個の対応関係から構成されている。商品分類についてはSITC・R2は「Commodity Indexes for the Standard International Trade Classification, Revision 2(Statistical Papers Series M no.38/Rev, Vol.I, United Nations 1981)」、SITC・R3は前述した「Standard International Trade Classification Revision 3」の各商品分類表から得られた詳細分類コードとして定義された4桁あるいは5桁コードである。5桁コードを持たないもののみが4桁コードで定義されている。それに対して基本モデルGRT₃₂で使用されている個別分類コードはSITC・R2については4、5桁の2種類、また、SITC・R3については3、4、5桁の3種類を含んでおり5桁以外は必ずしも詳細分類コードではなくそれらの上位分類コードが用いられていることもある。「GRT₃₂(IDE)およびアジア研統一国コード表」の表1-1にはGRT₃₂で使用されている詳細分類ではない個別分類コードとそれに対応する詳細分類コードが示されている。3桁の個別分類コードはSITC・R3の

746
747

の2個である。

基本モデルGRT₃₂で使用されている分類コードはすべてが詳細分類コードではないのでこの分類コードと商品分類表で用いられているSITC・R3およびSITC・R2の詳細分類コードとの一致性が問題となる。SITC・R3については商品分類表は上位概念のコードまで含めたすべてのコードを持ち、レコード数にして4,194である。これから詳細分類である4あるいは5桁コードを取り出すと3,121となる。

GRT₃₂は5,245の対応関係からなっており、この中からSITC・R3の重複無しのコードは2654である。前述したようにこのコードは上位概念のコードを203含むのでこれを取り去り、その代わりに詳細分類コード670を加えると3,121となり商品分類表の詳細分類コードの個数と一致する。SITC・R2については商品分類表は上位概念のコードまで含めたすべてのコード数は2,257である。これから詳細分類である4あるいは5桁コードを取り出すと1,831が得られる。このコードは上位概念のコードを2個含むのでこれを取り去り、その代わりに詳細分類コード3を加えると1,832となり商品分類表の詳細分類コードの個数と一致する。

前述した方法を適用して基本モデルGRT₃₂の対応関係をグループ化してそのグループをタイプ分けをする。その結果の要約が表3と表4に示されている。

表3 各対応関係コード表モデルのタイプ別グループ数

タイプ 種類	total		type					
	1	2	3	4	4A	4B		
GRT ₃₂ (%)	1007 100.0	795 78.9	9 0.9	115 11.4	88 8.8	65 6.5	23 2.3	
GRT ₃₂ (IDE) (%)	1322 100.0	975 73.8	15 1.1	201 15.2	131 9.9	75 5.7	56 4.2	

(出所) 著者作成 (注) 上の数字はグループの個数、下の数字はそのパーセントを示す。

表4 各対応関係コード表モデルのタイプ別個数

タイプ 種類	total		type					
	1	2	3	4	4A	4B		
GRT ₃₂ (%)	5245 100	795 15.2	18 0.3	303 5.8	4129 78.7	408 7.8	3721 70.9	
GRT ₃₂ (IDE) (%)	4306 100	975 22.6	30 0.7	581 13.5	2719 63.1	411 9.5	2309 53.6	

(出所) 著者作成 (注) 上の数字は対応関係の個数、下の数字はそのパーセントを示す。

表3から基本モデルGRT₃₂を構成するグループの個数はタイプ1は795、タイプ2は9、タイプ3は115個からなりこの3種類でグループ全体の91.2%を占める。この3種類を表4により対応関係の個数で見ると1,114でありこの数は全体の21.2%しか占めない。すなわち、GRT₃₂の対応関係コード表の

対応関係の78.8%はタイプ4からなり、しかもその中で一番処理がやっかいなタイプ4Bに属しているグループの個数は23でありながら対応関係の個数は3,718と圧倒的に多く、対応関係コード表の約70%を占めている。

また表5にはタイプごとにグループに含まれる対応関係の個数が示されている。グループに含まれる最大の個数は2,766であり、この中には先頭の1桁目をとっても0から9までのすべてのコードが存在しており、グループのままでは対象とする範囲が広すぎてグループの特徴が明らかにできない。また、対応関係を個別に検討していくと小さなグループであっても切断の必要と思われる箇所がいくつかでてくる。したがって、切断によるサブグループ化の必要が生じてくる。このような必要に応じて作成されたアジア経済研究所独自の切断によるモデルGRT₃₂(IDE)は次節以降で紹介する。

表5 GRT₃₂のタイプごとのグループに含まれるレコード数

要素の 個数	total		type					
			1	2	3	4	4A	4B
1	795	795	0	0	0	0	0	0
2	91	0	9	82	0	0	0	0
3	30	0	0	14	16	16	0	0
4	26	0	0	8	18	16	2	2
5	14	0	0	3	11	9	2	2
6	15	0	0	7	8	5	3	3
7	6	0	0	0	6	6	0	0
8	7	0	0	1	6	4	2	2
9	1	0	0	0	1	1	0	0
10	3	0	0	0	3	3	0	0
11	1	0	0	0	1	0	1	1
12	1	0	0	0	1	0	1	1
13	1	0	0	0	1	1	0	0
14	2	0	0	0	2	1	1	1
15	1	0	0	0	1	0	1	1
16	1	0	0	0	1	0	1	1
17	1	0	0	0	1	1	0	0
18	2	0	0	0	2	0	2	2
24	1	0	0	0	1	0	1	1
26	1	0	0	0	1	1	0	0
33	1	0	0	0	1	0	1	1
34	1	0	0	0	1	0	1	1
38	1	0	0	0	1	1	0	0
44	1	0	0	0	1	0	1	1
270	1	0	0	0	1	0	1	1
391	1	0	0	0	1	0	1	1
2766	1	0	0	0	1	0	1	1
total	1007	795	9	115	88	65	23	

(出所) 著者作成

第4節 アジ研モデルGRT₃₂(IDE)における「切断」の考え方

SITC・R3とSITC・R2を対応づけた基本モデルのうち、タイプ1、タイプ2およびタイプ3の対応関係のグループは、統計値の包含域が完全に一致するので両体系で分類された統計値を時系列で接続させても断層はおきない。ここでは、タイプ4に属するグループに「切断」の考え方をとり入れて、できるだけ統計値の連続性を保ちつつ、なおかつ構成要素の少ないシンプルなサブグループ化への分割を検討する。サブグループ化の目的は、2つあり、(1)グループの構成要素が、たとえばグループ888の2,766個のように多いと、グループ内に共通して存在している特性が見えてこない。この2,766個の要素から「結びつきの希薄な要素」を「切断」という方法で取り除くことによって、特性が見えるまとまりのあるグループにすることである。(2)グループの構成要素があまりに多すぎると貿易統計の利用面での制約となるため、統計値の連続性に大きな影響を及ぼさないような要素を「切断」することによって1つのグループを分割してシンプル化をはかることである。

SITC・R2からSITC・R3への改訂には、分類の基準の変更があることはすでに第4章で述べた。この変更のあった分類の基準を類型化すると、ほぼつぎのようにまとめられよう。

(1) 商品の原材料を分類基準にとり入れるか否か。原則的に原材料は最終製品の分類の基準である。繊維製(綿、羊毛、人造繊維等)、木製、皮革製、紙製、プラスチック製、金属製(鉄、銅、アルミニウム、貴金属、その他)、ガラス製などで、分類コードは同一にはならない。特に、CCCN-SITC・R2は原材料重視の分類体系である。しかしHS-SITC・R3はそれまでの過度な素材重視を一部見直し、同様な機能をもつ商品に1つの分類コードをもたせている。例としては、ランプその他の照明器具類(第4章第5節5-3(12)参照)、携帯用時計のバンド及びブレスレット並びにこれらの部分品(第4章第5節5-3(14)参照)がある。

(2) 駆動方式、機能、精度に大きな影響を与える方式を分類基準にとり入れるか否か。貿易商品分類ではこれらの方式が異なれば、分類コードも異なっていることが多い。しかし、商品によって方式を分類基準にとり入れる場合とそうではない場合がある。特に、SITC・R2ではとり入れていたが、SITC・R3では用途や機能を重視して、とり入れるのをやめた商品がある。分類体系の接続に影響がある

SITC・R2ではとり入れていたがSITC・R3ではとりやめた商品の例には、分析機器類、各種計器、自動制御機器類（これらの商品は、SITC・R2では、電気式、電子式、その他の方式、によって分類コードが別であったが、SITC・R3では方式にかかわらず、機能重視の分類に変更された）がある。

（３）用途を分類基準にとり入れるか否か。用途が分類の基準である場合は多いが、原材料の段階で用途が特定できない商品も多い。食用・飼料用・採油用・その他を、農産物の時点で決められない場合もあるだろう。ここでは、SITC・R2では「パルプ用」と用途限定があった木材の分類基準がSITC・R3ではなくなった例をあげる。パルプ用材及び木材（第４章第５節５－３（３）参照）。

また、R2では同一の分類コードに分類されていた「とうもろこし」のうちスイートコーンが、SITC・R3では穀物ではなく野菜とみなされ、分類コードが別になった。

SITC・R2では、いくつかの商品から構成される同一目的に使用されるセット物に適切な分類コードがなかったが、SITC・R3ではこうしたセットが１単位となっている商品に分類コードが新設された。例えば、トラベルセット、小売用の手道具または手工具のセット（箱に入った大工道具など）がある。

電気機器はSITC・R2でもSITC・R3でも家庭用か産業用で分類番号は原則的に異なる。ただし、家庭用か産業用（業務用）か区別が必ずしも明確でないため、SITC・R2の１つの分類コードに含まれる商品群とSITC・R3の１つの分類コードに含まれる商品群が完全に同一でない場合がある。例として、冷蔵庫及び冷凍庫等がある。

（４）処理・加工を分類基準にとり入れるか否か。原則的に、処理・加工が異なれば、分類番号も異なる。しかしSITC・R3で新たに分類基準となった商品には、反物状のキルティングした商品（第４章第５節５－３（７）－e参照）がある。

（５）サイズ、形、長さ及び１個の重量や容量など、分類規格の変更など。これらに該当する例として、鉄鋼の１次製品では、SITC・R2とSITC・R3で随所に分類基準の変更があった。鉄鋼及び鉄鋼製品（第４章第５節５－３（９）参照）。また、木材の例では、厚板と薄板の区分線（厚板と薄板の区分線が、SITC・R2の５ミリメートルからSITC・R3では６ミリメートルへ）が変更になった。

（６）使用する人（男性、女性、乳幼児等）を分類基準に取り入れるか否か。SITC・R2もSITC・R3も、衣類の分類で男性用のものは独立の分類コードが付されている。しかし、女性用と乳幼児用の衣類

の分類が、SITC・R2とSITC・R3では異なっている。すなわちSITC・R3では女性用と乳幼児用の衣類及びその付属品（第４章第５節５－３（７）－g参照）がそれぞれ分離独立した。

（７）SITC・R2作成の時点では存在しなかったため、適切な分類コードが付与されていない商品。これらの例として、光ファイバー、光ファイバーケーブル、人工衛星等がある。

第５節 アジ研モデルGRT₃₂(IDE)における「切断」の方法

SITC・R2とSITC・R3の対応関係コード表では第４節で説明したような分類基準の変更や新設が複雑にからみあってグループ４に属するような関係を多く生み出している。しかし、複雑にからみあった関係もグループを構成する個々の要素を検討すると、SITC・R3とSITC・R2のコードの結びつきには強弱がある場合がほとんどである（注４）。１グループの構成要素が等しく弱い結びつきから構成されているグループもあるが、こうした場合の処理は後述する。まず、結びつきの強弱を決める関係を整理してみよう。（１）包含される商品の範囲が完全一致、（２）１項目コードに包含される個別商品が一部分のみ対応（対応する比率はさまざま）、（３）商品の原材料が同一、またはかなり関係があるかあるいは一部分関係がある、（４）機能上あるいは用途上関係がある、（５）製造工程や加工面で関係がある、（６）具体的商品ではさしあたっては対応関係はなさそうだが、対応関係を有する商品が想定される、などにまとめられる。

このうち、（１）は問題なく強い結びつきであり、（２）から（５）は各要素によって強い結びつきからごく弱い結びつきまでさまざまである。（６）は弱い結びつきとみる。サブグループ化のための「切断」の候補の対象となる要素はまず（６）であり、（２）から（５）の関係で接続されている要素については結びつきの弱い要素である。これらの候補の中からサブグループ化の要因となる要素のみが「切断」の要素となる。

「切断」の候補とする過程で第４節にあげた２つの目的のうちより重視するのは、特性が見えるグループにサブグループ化することである。逆にグループの構成要素の数が非常に多くてもグループの特性が明確であるならば、その中にグループの特性を乱さない結びつきの関係が弱い要素が入っていても「切断」の候補としない。

例としては、グループ604の「鉄鋼製品」がある。

これらの商品は大幅な分類の定義の変更により型材がワイヤーと結びついたり、フラットロール製品が棒鋼と結びついたりしている。これらのSITC・R3とSITC・R2の各コードは、多くのコードが10を越える数の腕で相互に結びつき、要素数は391個もある。こうしたグループにおいて「鉄鋼製品」より下位のグループの特徴でサブグループ化するための「切断」の候補を選び出すことは困難で、「切断」を強行すると非常に多くの要素が「切断」の候補の対象となって落ちてしまう数値が多くなることが想定され、SITC・R2による統計値とSITC・R3の統計値の連続性を乱す。したがって、「鉄鋼製品」というグループ604では要素の「切断」はおこなっていない。

強い結びつきの要素と弱い結びつきの要素から構成されるグループの中から、弱い結びつきの要素を「切断」で除去してサブグループ化した例をあげる。グループ777から、R3:775.21（家庭用冷蔵庫及び冷凍冷蔵庫）とR2:741.41（非家庭用冷蔵庫及び冷凍機、冷凍庫等）、R3:775.22（家庭用冷凍庫）とR2:741.41（同上）、の2つの要素を弱い結びつきとみなし、このグループから取り除く。この結果、グループ777は切断の要素の集まりと3個のサブグループに分かれる。この結果は表6に示されている。

表6 グループ「777」における切断とサブグループ化の具体例

[切断の要素]	775.21	741.41
[切断の要素]	775.22	741.41
[サブグループ1]	741.43	741.41
	741.45	
[サブグループ2]	775.21	775.21
[サブグループ3]	775.22	775.22

（出所）著者作成

ここで、左側のコードはSITC・R3、右のそれはSITC・R2を表し、R3:741.43（非家庭用冷蔵・冷凍キャビネット、ショーケース等）、R3:741.45（非家庭用または冷凍用の機器及びヒートポンプ）、R3:775.21（家庭用冷蔵庫及び冷凍冷蔵庫）、R2:775.21（家庭用冷蔵庫及び冷凍冷蔵庫）、R3:775.22（家庭用冷凍庫）、R2:775.22（家庭用冷凍庫）である。

グループ888のサブグループ化には最も苦慮した。このグループは、1桁目のコードがSITC・R3では「3」と「9」を除く「0」から「8」までのsection、SITC・R2では「0」から「9」までのすべてのsectionの2,766個の要素から構成される特性が全

く見えないグループである。このグループは要素数が非常に多いため、特性の見えるグループにするためにかなり大胆な「切断」を行わざるをえなかった。その結果、繊維糸や織物類もある程度原料別にサブグループ化できたし、衣類、プラスチック関係品、ゴム製品などもそれぞれ1つずつのグループにまとめることができた。しかし商品によっては「切断」の基準は必ずしも一定ではない。繊維糸や織物類については、より「切断」の基準を緩やかにして繊維糸と織物類をそれぞれ1つのグループとする考え方もあると思う。

一方プラスチック関係品では、「プラスチック1次製品」と「プラスチックの半製品」および「プラスチックの製品」の3グループに分けたかったが、SITC・R3とSITC・R2のこれらの商品のコードが相互に関係しあっていて、結局プラスチック関係品以上のサブグループ化は、接続の抽象モデルではマイナス面が大きいと考えておこなっていない。

衣類についても、貿易統計の最も多い利用のされ方も考慮にいれて、「切断」によって落ちるかもしれない統計値が多く出ることを避け、「衣類」としてのグループでの利用を選んだ。

国際貿易上さして重要ではない商品で、なおかつSITC・R3の1コードが多数のSITC・R2のコードとほぼ等しい弱さで結びついている商品については、接続の核になるコードが特定できないため対応に苦慮した。こうした例は、HS-SITC・R3で新設された項目コードにみられることが多い。この場合の対応の第1は、要素数が10以下位ならば「切断」を行わず、全部の結びつきを生かすようにする方法である。例としては、R3:695.7「手道具または手工具のセット：695.2～695.5までの2以上の項目の製品を小売用のセットにしたもの」がある。

第2は、全ての要素を切断する方法である。該当例は以下のとおりである。（1）R2の項目コードでSITC・R3のどのコードとも接続しなかったコード。R2：657.4「ゴム糸を用いた紡織用繊維の織物類およびトリミング（メリヤス編みまたはクロセ編みのものを除く）」についてはSITC・R3の81個のコードと対応づけがなされているが、どの結びつきもきわめて弱い結びつきである。R2:696.07「卑金属製の柄」はHSでは廃止されたコードである。

（2）SITC・R3の項目コードでSITC・R2のどのコードからも接続しなかったコード。R3:654.94「綿製を除くガーゼ織物（細幅織物類を除く）」はガーゼ織物の主素材のもめんのものは別にコードを持っており、このコードはSITC・R2の20個のコードと対応づけがなされているが、どの結びつきもきわめ

て弱い結びつきである。R3:658.99「小売用に包装された織物と糸から成るセット」はHS-SITC・R3の新設項であり、SITC・R2の42個のコードと対応づけがなされているが、商品の性格や結びつきの内容の検討により、SITC・R2と接続させない方がよいと判断した。R3:813.99「ランプその他の照明器具の部分品のうち、ガラス製・プラスチック製でないもの」はHS-SITC・R3の新設項の中の部分品でありSITC・R2の21個のコードと対応づけがなされているが、上記R3:658.99と同様の理由によりSITC・R2と接続させていない。

第6節 アジ研モデルGRT₃₂(IDE)

基本モデルGRT₃₂を第5節の考え方にしたがって切断して作成したのがアジア経済研究所作成モデルGRT₃₂(IDE)である。切断の方法は基本モデルGRT₃₂の対応関係の中で比較的關係がないと考えられる対応関係を切断の候補とし、第3節で説明したようにグループへ戻してもグループの構成に変化がないものはグループへ戻しそうではないものを切断の要素とすることである。表7には第4節で説明した切断の候補に対する判定条件の(1)から(4)までの結果が示されている。ここで(3)と(3)'はそれぞれSITC・R3とSITC・R2の分類コードがいずれのサブグループにも属していないという状況を表している。元へ戻す候補は(2)である。

GRT₃₂(IDE)を作成するために第5節に従って切断の候補として選び出された対応関係の個数は977であったが、表7に示されているように38個が元に戻されて切断の要素は939個である。基本モデルの対応関係の個数が5,245であったので切断の部分を取り除いた残りの4,306個に対してグループ化がおこなわれサブグループが得られる。グループの888番目のものは基本モデルGRT₃₂の中で最大であり、2766個の対応関係から成る。これに切断をおこなうと245個のサブグループに分かれる。しかしながら604、888-229、901、888-216は依然としてグループ内に200個を越える対応関係を持つ(注4)。しかしながらこれらのグループは要素の数が多くにも関わらず商品の内容から分離が不可能であるためこれ以上の切断はおこなっていない。GRT₃₂(IDE)の結果の要約は表3と表4に示されている。基本モデルに対してこのモデルはグループにして315個増えしており、しかも、最大グループの個数は391である。また、表7には切断されたグループのサブグループの個数も表示されている(注5)。

表7 切断の必要なグループと切断候補の調整

グループ番号	(4)	(3)	(3)'	(2)	(1)	サブグループの個数
26	0	0	0	0	1	2
52	0	0	0	0	1	2
56	0	0	0	0	17	18
118	0	0	0	0	1	2
285	0	0	0	0	2	3
367	0	0	0	0	8	9
422	0	0	0	0	16	12
511	0	0	0	0	1	2
659	0	0	0	0	2	3
738	0	0	0	0	5	3
746	0	0	0	0	2	3
752	0	0	0	0	1	2
777	0	0	0	0	2	3
783	0	0	0	0	5	4
802	0	0	0	0	1	2
827	0	0	0	0	1	2
830	0	0	0	0	7	8
846	0	0	0	0	1	2
870	0	0	0	0	1	2
888	1	85	81	38	691	245
913	0	0	0	0	1	2
929	0	0	0	0	2	3
953	0	0	0	0	1	2
974	0	0	0	0	1	2
990	0	0	0	0	1	2
total	1	85	81	38	772	340

(出所) 著者作成

「GRT₃₂(IDE)およびアジ研統一国コード表」の表1-2に示されているのがGRT₃₂(IDE)であり、同一グループ内の切断の要素とサブグループの両者が含まれている。表1-2に表されたGRT₃₂(IDE)の各項目の記号とそれが表す内容は次の通りである。

G	グループの一連番号
SG	サブグループの一連番号
SG-T	サブグループのタイプ
R3-F	SITC・R3の頻度
R3-NM	SITC・R3の名称
R3	SITC・R3
R2	SITC・R2
R2-F	SITC・R2の頻度
R2-NM	SITC・R2の名称

切断の要素はSGがグループ内の0番で表される対応関係であり、SG-Tは空白のまま残されている。切断の要素である対応関係には、R2-FのところにSITC・R2の分類コードが所属しているサブグループの番号、R3-FにはSITC・R3の分類コードが所属しているサブグループの番号が表示されている。

サブグループはSGがグループ内の一連番号で表される対応関係の集まりである。R2-FはSITC・R2

の頻度、すなわち、結合のための手の個数である。R3-Fも同様である。

例として、第5節で取り上げたグループ777を用いてGRT₂₃(IDE)を表現する。このグループは2つの切断の要素と3つのサブグループから構成されている。分類AをSITC・R2、分類BをSITC・R3と置き換え、グループの一連番号は777であるのでGは777であり、切断の要素C₇₇₇はSGを0、AG₇₇₇(1)はSGを1、AG₇₇₇(2)はSGを2、AG₇₇₇(3)はSGを3として表される。SG-Tについては切断の要素は空白、SGが1であるサブグループはタイプ2なので2、SGが2および3であるサブグループはタイプ1なので1となる。

最初の切断の要素のR3:77521はサブグループの中のAG₇₇₇(2)に所属しているのでR3-FにはAG₇₇₇(2)のSGの番号である2が入る。同じく、切断の要素のR2:74141はサブグループの中のAG₇₇₇(1)に所属しているのでR2-FにはAG₇₇₇(1)のSGの番号である1が入る。2番目の切断の要素についても同じことがいえる。

サブグループはグループから切断の要素を取り除いたG₇₇₇-C₇₇₇に対してグループ化をおこなって作成されるので、R2-FとR3-Fはサブグループの中だけで有効な結合の手の個数である。R3:77521は切断の要素に含まれているが、AG₇₇₇(2)の中では結合の手の個数は1個であるので、R3:77521に対応するR3-Fは1となる。以上をまとめて、GRT₃₂(IDE)は図2のように表される。

グループ内に切断の要素が存在しない場合には、グループがサブグループに分かれないので、SGはすべて1で表される。

R2-NMとR3-NMの項で表される商品分類の名称は30桁に縮小されているが、詳細について知りたい場合には第3節で紹介した各商品分類表を参照すると良い。この表からGのみの対応関係を見たとき、この対応関係は切断を考慮していないので基本モデルGRT₃₂になる。

おわりに

商品分類SITC・R3とSITC・R2の対応関係コード表の基本モデルGRT₃₂に対してグループおよび切断によるサブグループで変換するという方法を適用してアジア経済研究所の独自の切断により作成したモデルがGRT₃₂(IDE)である。これらのモデルGRT₃₂、GRT₃₂(IDE)は対応関係コード表からのみ得られた抽象モデルである。抽象モデルに対して、データにもとづいた実際の商品分類の変換にさいしては各国の貿易取り引き状況に応じて取り引きの無い対応関係から切断が生じて抽象モデルよりもさらに細分化されたグループやサブグループを持つ具体モデルができあがる(補論の「D. 抽象モデルと具体モデル」を参照のこと)。さらに、取り引き額の大きさを考慮したさらなる切断モデルも考えられる。すなわち、ある基準値よりも小さい取り引き額を持つ個別分類コードは取り引きがないとみなしてグループ化をおこなうということである。

本章でおこなった切断の試みはまだ第1段階であり、今後に残されている課題がいくつかある。

(1)「切断」の評価基準を検討し、切断の要素を再調整をする。

図2 GRT₃₂(IDE)の具体例(グループ777)

G	SG	SG-T	R3-F	R3-NM	R3	R2	R2-F	R2-NM
777	0		2	household type refrigerators	77521	74141	1	non-domestic refrigerators, nes
777	0		3	household type deep-freezers	77522	74141	1	non-domestic refrigerators, nes
777	1	3	1	non-domestic refrigerators, nes	74143	74141	2	non-domestic refrigerators, nes
777	1	3	1	other refrigerators equip	74145	74141	2	non-domestic refrigerators, nes
777	2	1	1	household type refrigerators	77521	77521	1	domestic refrigerators
777	3	1	1	household type deep-freezers	77522	77522	1	household type deep-freezers

(出所) 著者作成

(2) 本来同一のグループに含まれるべきであるのに、たまたまSITC・R2とSITC・R3の間で対応関係が完結して別のグループになってしまう商品グループが発生する。これらの商品グループについてはできるだけ同一の範囲の商品グループどうしを接続させるという趣旨からグループ間の統合が必要になる。

(3) SITC・R2とSITC・R1の対応表では、3桁間、2桁間、1桁間の利用についても斉合関係を保持させているが、GRT₃₂(IDE)ではこの問題を解決していない。貿易統計をアグリゲートして利用したい要求にどのようなツールを提供できるのか、階層構造をも含めたグループ間の再構成が必要である。

また、本章では切断を中心としたサブグループによる枠作りが中心になってしまったが、グループ化の最終目標としては体系の異なる2つの両商品分類を共通した概念で体系化した「アジア経済研究所作成の共通した商品分類の利用ガイド」のようなものを完成させることである。時間の制約上本章ではそこまで到達できなかったが、2つの両商品分類を共通した概念で体系することやその最適な検索方法については今後の検討課題に残して置くことにする。

(注1) 本章のグループ化は第7章の定理3を考え方基礎としている。データベース内に分類階層が用意されており、その中に両分類が定義されているなら、この両分類から共通に導出可能な最も詳細な分類FCD(Finest Common Derivative)を機械的に求めることができる。

(注2) 1個の個別分類コードに複数個の分類コードが対応しているとき、この対応関係に統計量を割り振るにはウェイトを付けて複数に配分することが考えられるが、これを配分ウェイトによる方法と呼ぶ。説明は省略するが、タイプ4には配分ウェイトが代数方程式により一意的な解を持つタイプ4aと一意的には解を持たないタイプ4bに分けることができる。

(注3) 結合の手の多い分類コードを切断するには、グループに属する対応関係の中から分類Aに属する分類コードの結合の手の数が m_1 個以上、SITC・R2の結合の手の数が m_2 個以上のものを取り除いた対応関係の集まりに対してグループ化を試みる。そのとき得られたグループの個数を $gp(m_1, m_2)$ で表す。 m_1 と m_2 のそれぞれにグループ内の最大の値からはじめて数を減らしながら与え、 $gp(m_1, m_2)$ の値の状

態を確かめる。局所的に $gp(m_1, m_2)$ が最大の値を持つような m_1 と m_2 を取り除きの基準とする。そのさい取り除いた対応関係が少なく、同時にグループの個数が多くなるような組み合わせを最適な選択とする。少数個の組み合わせの切断は m 個の中から n 個を取る組み合わせの数だけ計算をしなければならないのでグループの要素が多くなると膨大な計算時間を必要とする。

(注4) 第888番目のグループでサブグループが第229番目を888-229と表す。

(注5) 表7のサブグループの個数の合計は340となっているが、グループの個数として既に25あるのでサブグループ化によって増加したグループの個数は340から25を引いた315個となる。また、表5のGRT₃₂(IDE)のグループ数1,322からGRT₃₂のそのの1,007を引いても同じ結果を得ることができる

【参考文献】

[1] United Nations, Standard International Trade Classification, Revision3, ST/ESA/STAT/SER.M/34/Rev.3., New York, 1986.

[2] 佐藤英人「要約データの基礎概念とデータベース内での推論—世界貿易統計データベースを例として—」(木下宗七、野田容助編『世界貿易データシステムの整備と利用』統計資料シリーズno.67、アジア経済研究所統計調査部、1995年2月)

[3] 山本泰子「貿易統計における商品の分類」(木下宗七、野田容助編『世界貿易データシステムの整備と利用』統計資料シリーズno.67、アジア経済研究所統計調査部、1995年2月)

(補論)

A. グループ化の方法について

本章のグループ化は第7章の定理3のFCD(Finest Common Derivative)を求めることである。このFCDの作成を合成された類別関数、

$$(f^{-1} \circ g \circ g^{-1} \circ f)^n \quad n=1 \dots$$

を用いずに、データ処理という立場からまとめてみる。処理を簡単にするためにグループ化したい対応関係コード表に対して、分類Aと分類Bごとに分類コードの一連番号を付ける。次に分類A、分類Bの順にこの一連番号を昇順に並べる。このようにして作られた対応関係コード表の例を表A1に示す。グループ化は次のようにおこなう。

(1) 分類Aで同一番号を持っているものは同一グ

ループのメンバーである。同一グループに属している分類Bの番号のすべてをグループを識別する共通の識別記号で置き換える。この共通識別記号はグループ内に存在する最初分類Bの番号で代替する。

(2) 分類Aのすべての分類コードについて(1)の処理を繰り返す。

(3) グループごとに決められた識別記号を基準に並べ替え、識別記号に一連番号を付ける。この識別番号、または、一連番号でまとめられたのがグループである。

表A1 対応関係コード表の例

コード番号	分類		一連番号 A B	分類Bの変更 (1) (2) (3) (4)			
	A	B					
1	51561	51562	1 2	2	2	2	2
2	51561	51563	1 3	2	2	2	2
3	51561	51564	1 4	4	2	2	2
4	51562	51561	2 1	1	1	1	1
5	51562	51565	2 5	5	5	1	1
6	51562	51566	2 6	6	6	6	1
7	51563	51563	3 3	2	2	2	2
8	51564	51562	4 2	2	2	2	2
9	51564	51564	4 4	4	2	2	2
10	51565	51563	5 3	2	2	2	2
11	51565	51564	5 4	4	2	2	2
12	51566	51567	6 7	7	7	7	7
13	51567	51568	7 8	8	8	8	8
14	51568	51568	8 8	8	8	8	8
15	51569	51562	9 2	8	2	2	2

(出所) 著者作成

表A1にある対応関係コード表について実際にこの処理をおこなってみる。最初のレコードの分類Aは連番1であり、これに分類Bの連番2が対応している。分類Aの2番目は同じく連番1であり、それに分類Bの3が対応している。この2つのレコードは分類A共通なので同一グループに属しているが、分類Bとの関わりで同じグループのメンバーになるためには同一の識別記号が必要になる。分類Aの連番1には既に分類Bの連番2が付けられているので連番の2を識別記号とみなして分類Bの連番3のすべてを連番2に置き換えることで分類Bに同一識別記号を与える。表の「分類Bの変更」の(1)にその結果が示されている。分類Aの3番目は同じく連番1であり、それに分類Bの4が対応しており、連番の4はこの同一グループのメンバーである。分類Aの連番1に属すグループには既に分類Bの連番2が付けられているので表の「分類Bの変更」の(2)のように分類Bの連番4のすべてを連番2に置き換え同一識別記号を与える。

分類Aの4番目は連番2であり、それに分類Bの1が対応している。分類Aに関しては異なったグル

ープなのでそのままにしておく。分類Aの5番目は同じく連番2であり、それに分類Bの5が対応している。分類Aの連番2には既に分類Bの連番1が付けられているので分類Bの連番5のすべてを連番1に置き換える。

表A1の分類Bの(3)にその結果が示されている。この処理を最後まで続ける。分類Bが同じ連番を持っていればそれは同一グループのメンバーになる。この例では結果は次のように4つのグループに分かれることになる。レコード番号でグループを表現すれば、

$$G_1 = \{4,5,6\}$$

$$G_2 = \{1,2,3,7,8,9,10,11,15\}$$

$$G_3 = \{12\}$$

$$G_4 = \{13,14\}$$

である。また、この4つのグループは次節で述べるグループのタイプ分けをするとタイプ2、タイプ4、タイプ1、タイプ3にそれぞれ分類される。特にすべてが同じ番号の1になれば1つのグループしか存在しないことを表す。

B. グループ化された対応関係の関連情報

第1節の対応関係のグループ化とそのグループのタイプ分けの情報は表B1に示したように対応関係コード表に付加される。先頭から4桁はグループの一連番号、5から8桁には次節で説明するサブグループの一連番号、10から11桁は対応関係のタイプの種類、13から22桁には分類Aのコード(1行目の例は00122である)、23から32桁は分類Bのコード(1行目の例は00122である)、33から37桁は分類Aのコードの頻度(1行目の例は00122の頻度の1である)、38から42桁は分類Bのコードの頻度(1行目の例は00122の頻度の2である)、43から47桁は分類Aのコードの同一グループ内の一連番号、48から52桁は分類Bのコードの同一グループ内の一連番号を表す。

表B1 グループ化された対応関係コード表とその関連情報

....*.1....*.2....*.3....*.4....*.5..											
1	1	4a	00122	00122	1	2	1	1			
1	1	4a	01112	00122	2	2	2	1			
1	1	4a	01112	01112	2	1	2	2			
2	1	1	01113	01113	1	1	1	1			
3	1	3	01114	01115	1	2	1	1			
3	1	3	01115	01115	1	2	2	1			

(出所) 筆者作成

(注) 先頭のスケールはデータには付いていないが、桁数を表すために表示している。

この例では6個ある対応関係が3つのグループに分かれている。最初のグループの対応関係はタイプ4Aであり、その要素の数は3個である。2番目のグループはタイプ1でその要素数は1個、3番目のグループはタイプ3で要素数は2個である。

C. サブグループ化された対応関係の関連情報

対応関係のグループ化とそのグループのタイプ分けの情報は前述したように表3に示される。この例ではサブグループが存在しない例なので8桁目はすべて1になっていてグループとサブグループは同じ内容である。サブグループが存在する場合には表C1のように同一グループに対してサブグループの連番の8桁目がいくつかに分かれて表示される。この例は $n=2$ の場合である。

表C1 グループ化された対応関係コード表とその関連情報

....*1....*2....*3....*4....*5....
1	1 4a 00122	00122	1	2	1
1	1 4a 01112	00122	2	2	1
1	1 4a 01112	01112	2	1	2
1	2 3 01113	01115	1	2	1
1	2 3 01114	01115	1	2	2

(出所) 筆者作成

(注) 先頭のスケールはデータには付いていないが、桁数を表すために表示している。

D. 抽象モデルと具体モデル

基本モデル GRT_{32} の i 番目のグループを G_i 、取り除く切断の対応関係のグループを C_i で表す。 G_i 、 C_i に対してグループ化をおこないサブグループを作る。そのときに得られた n 個のサブグループを $SG_i(j)$ $j=1\cdots n$ とする。グループ G_i は、

$$G_i = \{ SG_i(1) \cup \cdots \cup SG_i(n_i) \cup C_i \} \cdots \cdots (1)$$

と分割される。分割なので、

$$SG_i(j) \cap SG_i(k) = \phi \quad j \neq k$$

$$SG_i(j) \cap C_i = \phi$$

である。グループ化によってもサブグループが生じない場合には $n=0$ であり、 $G_i=C_i$ となるが、これを1つのサブグループが生じたと解釈して $n=1$ 、 $G_i=SG_i(1)$ 、 $C_i=\phi$ とする。これからサブグループが生じないということと切断が生じないということは同じであることがわかる。このように対応関係コード表とその切断のみによって得られたモデルを抽象モデルという。もちろん切断なしの基本モデルも抽象モデルである。

抽象モデルに対して、データにもとずいた実際の商品分類の変換にさいしては、貿易取引状況に

応じて取り引きの無い対応関係や少額取引額を無視した対応関係から切断が生じて抽象モデルよりもさらに細分化されたグループやサブグループを持つモデルができる。このモデルを具体モデルとする。取引額 e 以下を取り引きのない対応関係とすれば具体モデルは e の関数として定義される。

具体モデルは抽象モデルに対する切断から生じるので、サブグループ $SG_i(j)$ はさらにいくつかのサブグループと切断に分割される。

$$SG_i(j) = \{ AG_{ij}(1) \cup \cdots \cup AG_{ij}(n_{ij}) \cup c_{ij} \} \cdots \cdots (2)$$

ここで、

$$AG_{ij}(m) \cap AG_{ij}(k) = \phi \quad m \neq k$$

$$AG_{ij}(k) \cap C_{ij} = \phi$$

である。この(2)式を(1)式に代入すれば、

$$G_i = \{ AG_{i1}(1) \cup \cdots \cup AG_{i1}(n_{i1}) \cup \cdots \cup AG_{in}(1) \cup \cdots \cup AG_{in}(n_{in}) \cup C_i^* \} \cdots \cdots (3)$$

となる。ここで、

$$C_i^* = \{ C_i \cup c_{i1} \cup \cdots \cup c_{in} \}$$

である。したがって、具体モデルは基本モデルをもとにして抽象モデルのための切断 C_i とデータから生じた切断 $\{c_{i1} \cdots c_{in}\}$ との両方を取り除いた $G_i-C_i^*$ をグループ化することによって得られる。切断が生じなければ具体モデルは抽象モデルに一致する。貿易の取引額は国、輸出入、年、相手国によって異なるため、同一の抽象モデルをもとにしても一般的に具体モデルは対象とする国、輸出入、年、相手国の組み合わせに応じて異なる。データ処理のさいには抽象モデルに対して貿易取引額に応じた切断をおこなったあとグループ化することで具体モデルを作成できる。

表D1に具体モデルを前提としたグループ化された対応関係とその関連情報が示されている。9桁から12桁にかけてサブグループをさらに分割した情報が示されている。

次に複数の国AとBを対象とすることを考える。抽象モデルのサブグループを $SG_i(j)$ とすると、これに対応するA国によって作られる具体モデルを $AG_{ij}(k)[A]$ 、そのときの切断を $c_{ij}(A)$ 、B国のそれを $AG_{ij}(k)[B]$ 、 $c_{ij}(B)$ とする。グループ化は切断によって決まるので

$$c_{ij}(A) \cup c_{ij}(B) = c_{ij}^{**}$$

となるように切断 c_{ij}^{**} を定める。そうすると $G_i-C_i^{**}$ はAとBとで共通になり、2国間の具体モデルはこれをグループ化することによって得られる。2国以上の場合にも同じことがいえる。データ処理の

表D1 具体モデルを前提としたグループ化された
対応関係コード表とその関連情報

...	*	...	1	...	*	...	2	...	*	...	3	...	*	...	4	...	*	...	5	...
1	1	1	1	00122				00122				1	2	1	1					
1	1	2	3	00123				00123				2	2	2	1					
1	1	2	3	01112				00123				2	2	2	1					
1	2	1	4a	01113				01114				1	2	1	1					
1	2	1	4a	01113				01115				1	2	1	1					
1	2	1	4a	01114				01115				1	2	1	1					
1	2	2	1	01116				01116				1	2	2	1					

(出所) 筆者作成

(注) 先頭のスケールはデータには付いていないが、
桁数を表すために示した。

さいには対象国である複数の国を込みにして1つの
国と見なし、抽象モデルに対して貿易取引額に応じ
た切断をおこなったあとグループ化することで複数
国の具体モデルを作成できる。